Socio-cognitive Architectures for Adaptable Autonomous Systems

ONR Program Code 34

August 2008

At a Glance

What is it?

- The Socio-cognitive architectures form the basis for software platforms to be installed on adaptable autonomous systems:
  - *Computationally instantiated theory of the entire cognitive system, including perception, cognition, learning and action, constrained by computational, neural, and psychological findings.*
  - *Socio-cognitive architecture focuses on cognition about other agents, including their beliefs, desires, intentions, and obligations.*

How does it work?

- Socio-cognitive architectures are able to reason directly about the mental states of other agents (either human or non-human) by maintaining "mental models" corresponding to the mental perspective of those agents.

What will it accomplish?

- Autonomous systems capable of enhanced coordination with both human and machine teammates.
- Intelligent systems capable of natural language interactions through dialogue.
- Systems that are capable of rapid learning via discernment of teacher's intent. Similarly, systems that teach more effectively by understanding student intent and comparing against skilled performance.

Points of Contact

Paul Bello

(703) 696-4318

paul.bello@navy.mill

Cognitive architectures have been an investment priority for the Office of Naval Research (ONR) for more than 20 years, encompassing many success stories. Most notably, cognitive architectures have been broadly applied to training and education in the development of intelligent tutoring technologies used at Department of Defense (DoD) dependent schools. Additionally, validated cognitive models of naval operator performance are currently being applied to human-computer interface issues as a cost-effective technique for exploring new concepts in naval displays.

Socio-cognitive architectures are outfitted with mechanisms corresponding to neural and psychological theories regarding the human ability to reason about the mental states (beliefs, desires, intentions, etc.) of other humans. The human capability to predict and explain the behavior of others in terms of beliefs, desires and intentions is sometimes called "mindreading." In general, mindreaders must be able to construct and maintain two "mental models:" one of the world as it is seen first-person, and one as it is seen from the perspective of the target agent to be predicted. What sorts of mental operations can be performed within or between these mental models remains an open issue for the psychology community, but the computational cognitive modeling community is currently moving to extend or adapt current cognitive architectures to be able to hold two or more such models in memory.

Research Challenges and Opportunities

- *Representation of mental states within cognitive architecture: mental simulation of other agents, bodies of conceptual knowledge about other agents, hybrid techniques, integrating learning with mental-state reasoning for rapid behavior prediction and explanation*
- *Integrating probabilistic and relational knowledge: real-time discernment of intentional action representing uncertainty about the mental states of others, disambiguation of spoken utterances in natural language dialogue*
- *Socially-mediated learning: self-reflection/diagnosis on task performance and mental simulation of alternatives in future performance, discernment of teacher intent during learning-by-imitation to facilitate quicker knowledge/skill acquisition, teacher discernment of student intent during performance to provide effective remediation*